Полная изоляция клиентов в облаке для сервисов без изоляции на примере DNS

Георгий Меликов
VK Cloud



Кто мы – VK Cloud

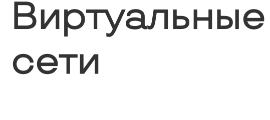
Инфраструктура (IaaS)

Виртуальные серверы

Гибкие конфигурации, неограниченное количество IP и безлимитный трафик в 1 ГБит/с.

Дисковые хранилища

Блочные (HDD, SSD) и объектные (S3). Классические хранилища или SDS (CEPH)



Виртуальные сети

Единая локальная сеть для серверов, приватный и публичный DNS, балансировка нагрузки и VPN.

Back-up и DR

Автоматическое восстановление IT-инфраструктуры в случае аварии

Платформа (PaaS)

Кластеры Kubernetes

Автоматическое масштабирование приложений и выстраивание быстрых DevOps-процессов.

Managed-базы данных

Полная автоматизация жизненного цикла работы СУБД: MySQL, PostgreSQL, MongoDB, Redis, Postgres Pro, ClickHouse, Arenadata DB, Tarantool Cloud

Cloud Big Data

Преднастроенные инструменты для хранения, обработки и анализа больших данных

Cloud ML Platform

Платформа для полного цикла ML-разработки и совместной работы Data-команд

Маркетплейс

Сервисы партнеров VK

Расширяющаяся экосистема решений на единой платформе

Сервисы VK

Прикладное ПО и программы для разработчиков





Infrastructure as a service

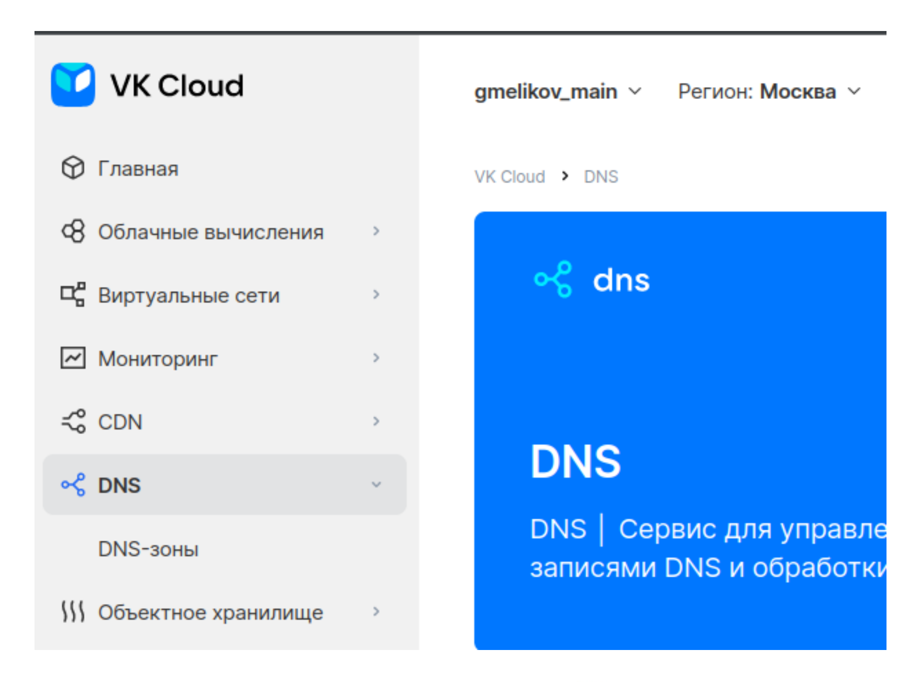
- Compute – виртуальная машина (CPU, RAM)
- Persistent storage – хранилище
- Сеть
  - VM interconnect
  - Доступ в Интернет





Infrastructure as a service

- Compute – виртуальная машина (CPU, RAM)
- Persistent storage – хранилище
- Сеть
  - VM interconnect
  - Доступ в Интернет
  - Public DNS authority server

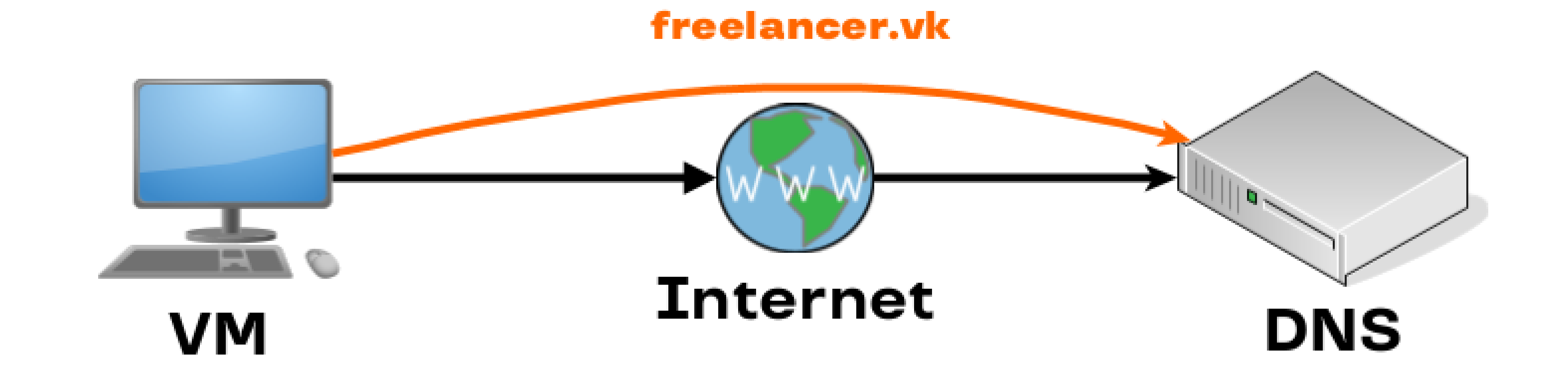






Domain Name System

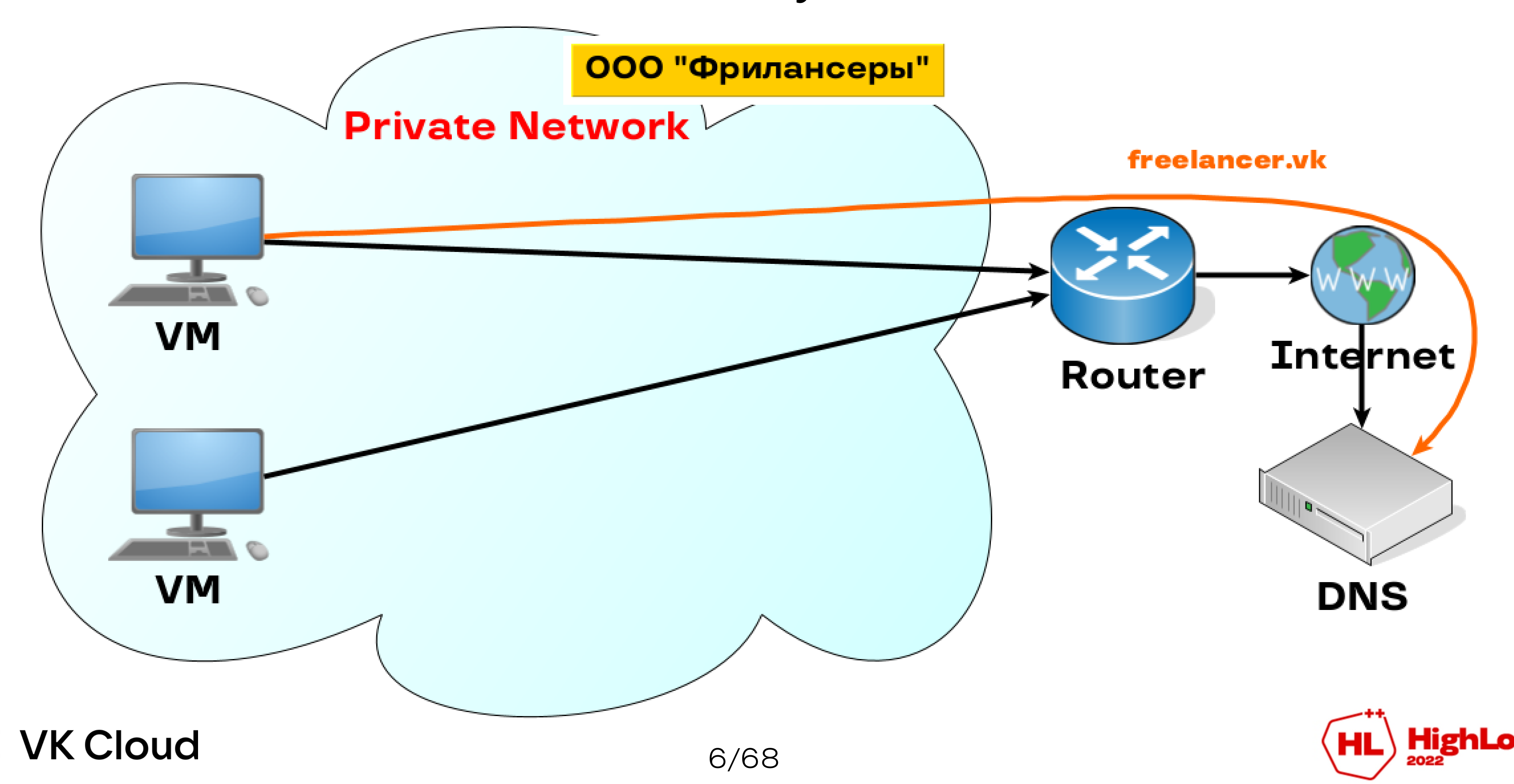
ООО "Фрилансеры"







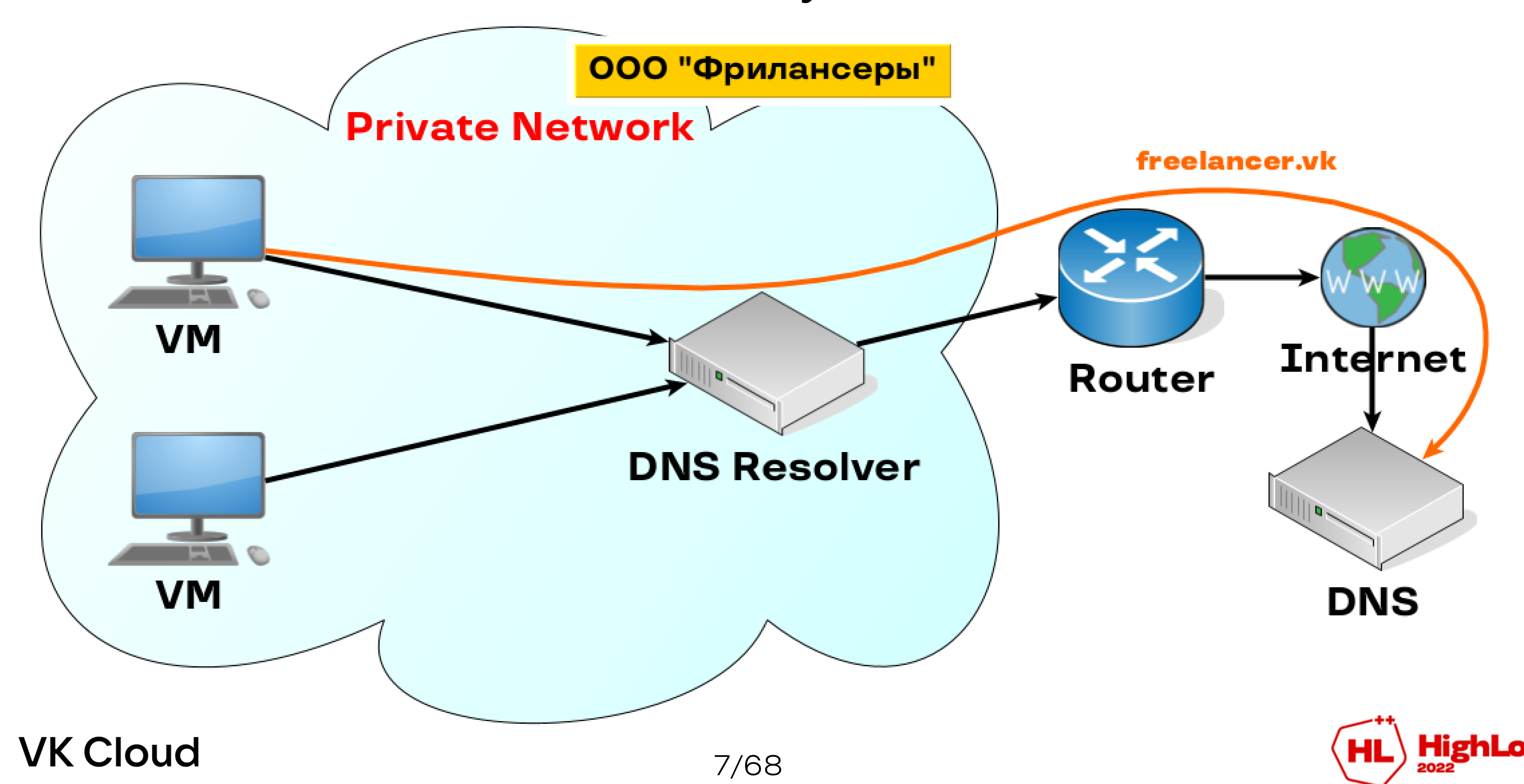
Private Domain Name System



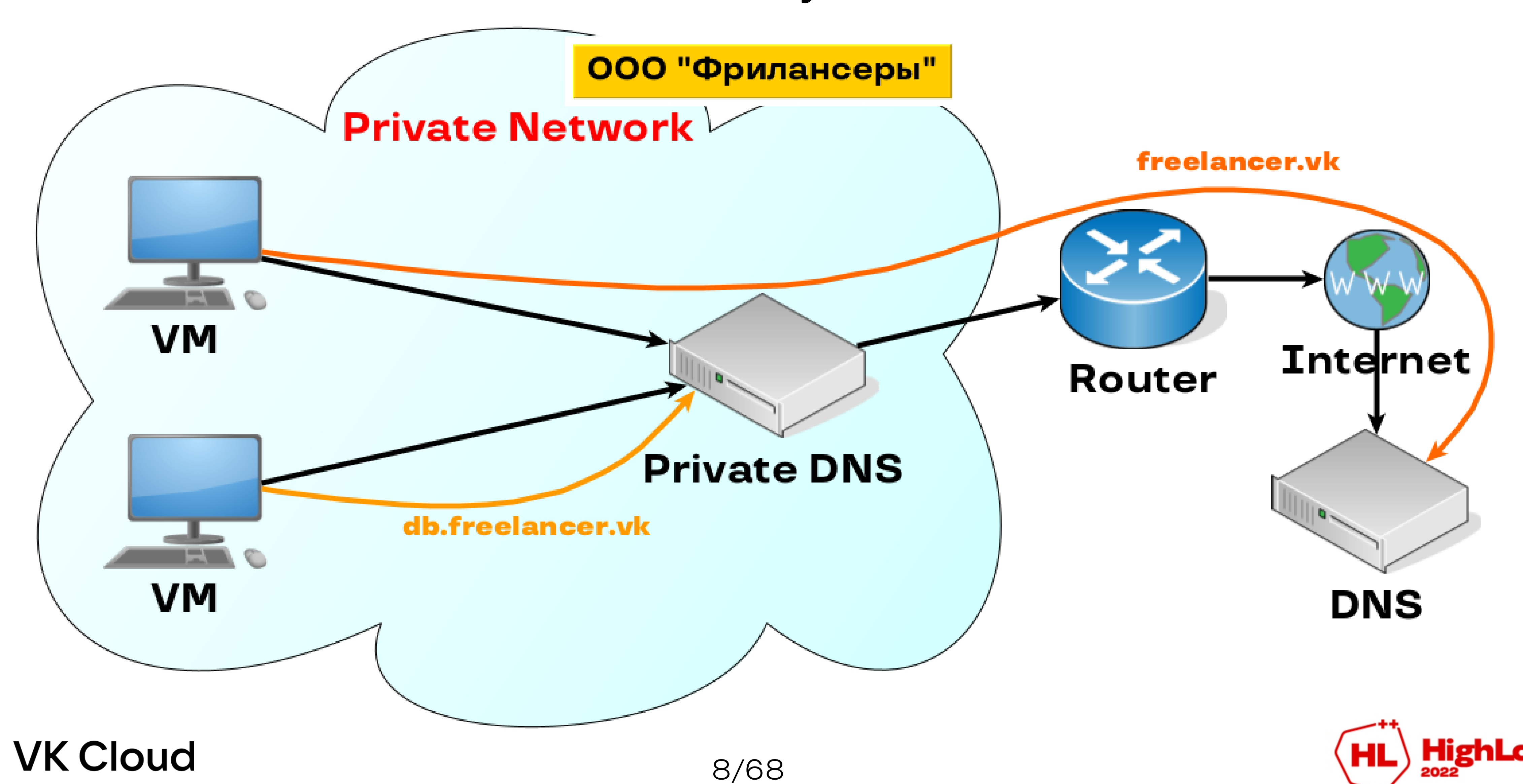
VK Cloud

HL HighLoad++ 2022

Private Domain Name System



Private Domain Name System



VK Cloud

HL HighLoad++ 2022

Private Domain Name System

ООО "Фрилансеры"

ООО "Аутсорсеры"





Разные ответы на одинаковый DNS-запрос

ООО "Фрилансеры"

```
Фрилансер:~$ dig freelancer.vk a
+short
10.0.0.1
```

ООО "Аутсорсеры"

```
Аутсорсер:~$ dig freelancer.vk a
+short
172.0.0.1
```







Service discovery через DNS

1. db.freelancers.local?

2. 10.0.0.1

3

VM

VM

DNS

10.0.0.1 (db1)

10.0.0.2 (db2)





Service discovery через DNS **без изоляции**

1. db.freelancers.local?

VM

VM

DNS

2. 1.2.3.4

3

10.0.0.1
(db1)

10.0.0.2
(db2)

1.2.3.4
(hacker-db)





Как сделать Private DNS as a service

- User API
- Authoritative DNS server







Наивный вариант

Project X
VM
Switch
DNS
DNS DB

Project Y
VM
Switch
DNS
DNS DB

Project Z
VM
Switch
DNS
DNS DB





Наивный вариант: проблема с количеством

Project X
VM
DNS
Switch
DNS DB

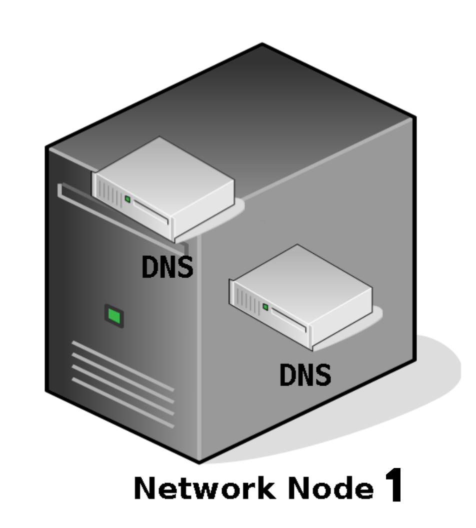
Project Y
VM
DNS
Switch
DNS DB

Project Z
VM
DNS
Switch
DNS DB



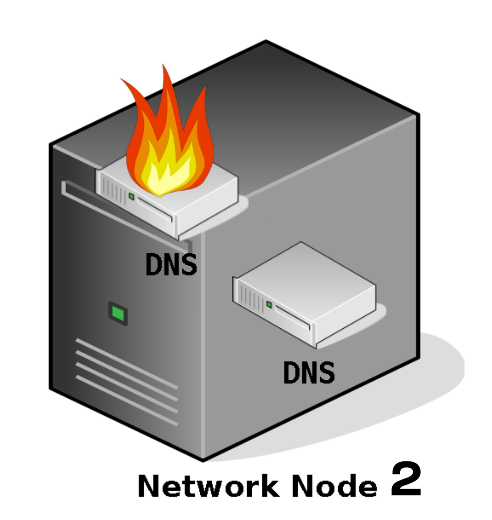


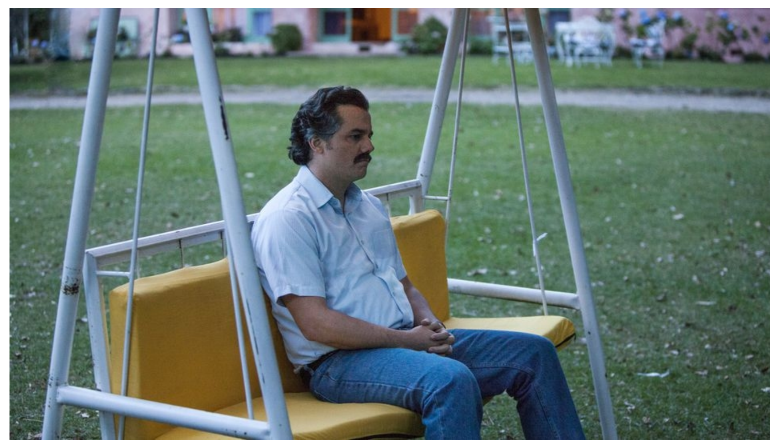
Наивный вариант: неравномерность



0-1RPS vs **1000+RPS**

99% **1%**









HL HighLoad++ 2022

Как сделать Private DNS as a service

- User API
- Authoritative DNS server

- Десятки тысяч проектов







Как работает Public DNS

Project X

VM VM

Project Y

VM VM

Project Z

VM VM

DNS

DNS DB





Один Dataplane на всех для **Private** DNS

```
Фрилансер:~$ dig freelancer.vk a
+short
10.0.0.1

Аутсорсер:~$ dig freelancer.vk a
+short
172.0.0.1
```



Project X

VM VM

Project Y

VM VM

Project Z

VM VM

DNS

DNS DB





Какой DNS-сервер выбрать?

Нужна мультитенантность

=

Фильтрация зон по проектам





Openstack Designate

- Мы начинали с OpenStack
- Нет мультитенантности как таковой
  - “Zones in Designate model the ownership concept from DNS itself, **where any given zone can only be owned by a single tenant.** However, while DNS is able to support a hierarchy of zones, there is **no support for delegating subzones to another tenant**, and one tenant cannot create zones that lie within the zone of another tenant.” @ https://docs.openstack.org/designate/latest/user/manage-zones.html





BIND

- Есть “views”
  - “One way to think about multiple views is as if each view were its own instance of BIND, with a fancy multiplexer in front of them all that determines which instance will receive each incoming request packet.” @ https://kb.isc.org/docs/aa-00851
  - “controversial feature”





PowerDNS

- Уже используем в нашем Public DNS





PowerDNS

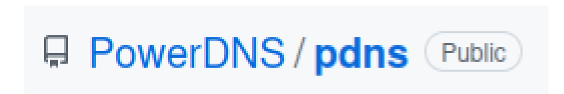
- Уже используем в нашем Public DNS
- “Views” нет
  - “Adding views would complicate the nameserver in many ways. Please run two copies of PowerDNS, they are both free!”
    @https://github.com/PowerDNS/pdns/issues/63

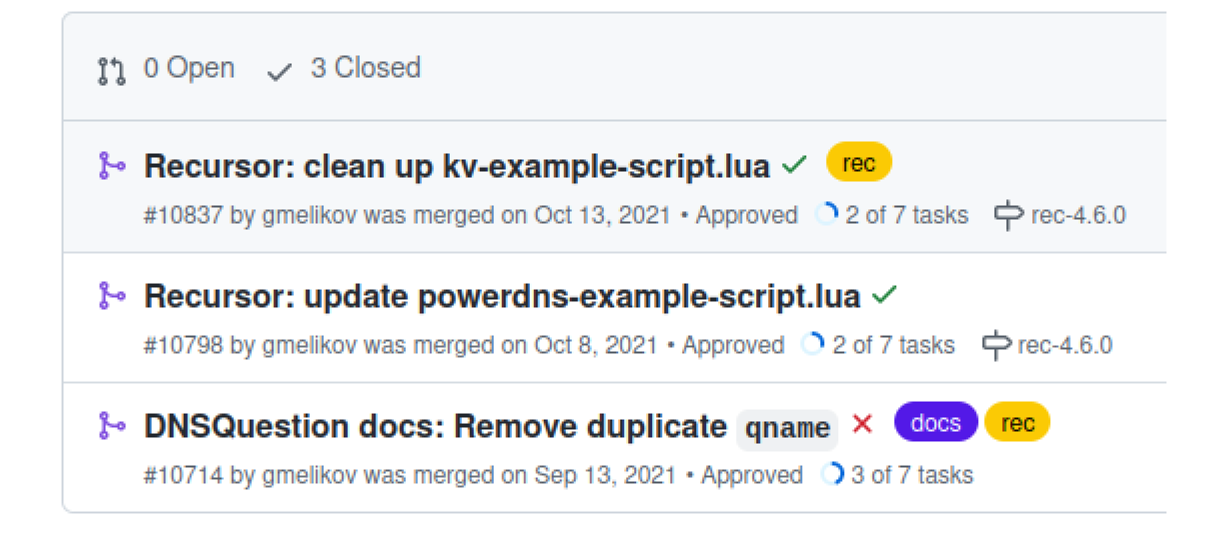




PowerDNS

- Есть LUA-плагины, изучим



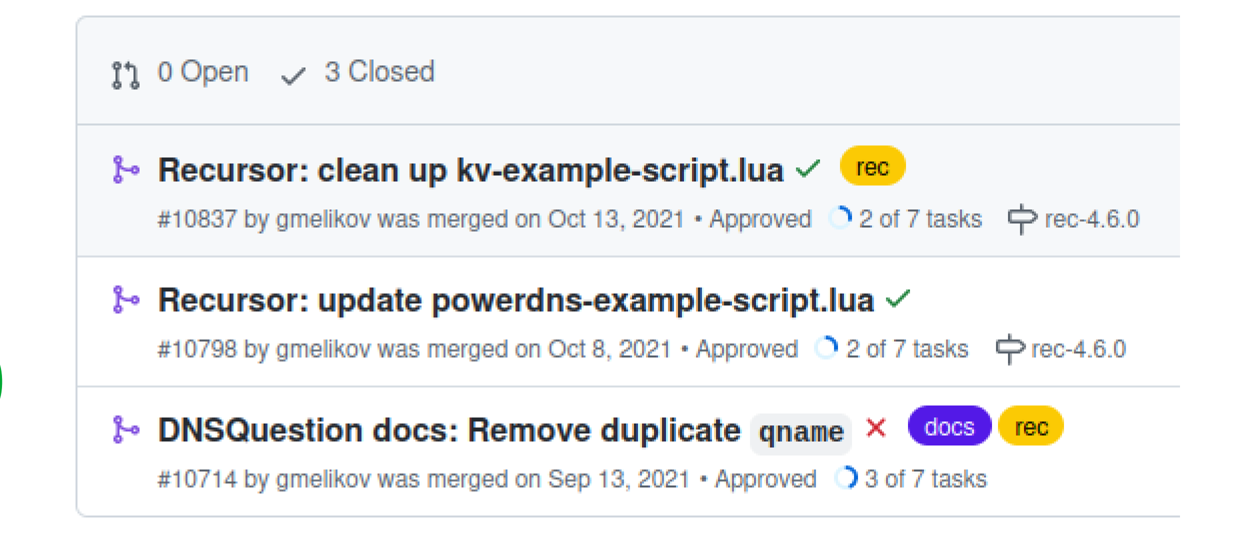






PowerDNS

- Есть LUA-плагины
- Не дают менять всё в запросах :(
- Стали контрибьюторами :)



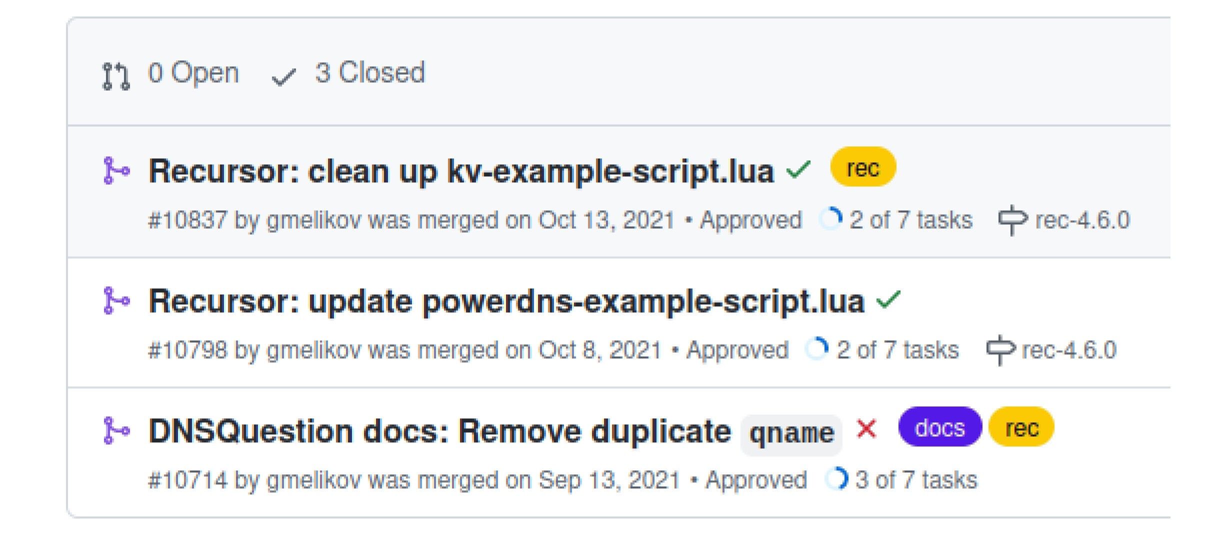




PowerDNS

- Есть LUA-плагины





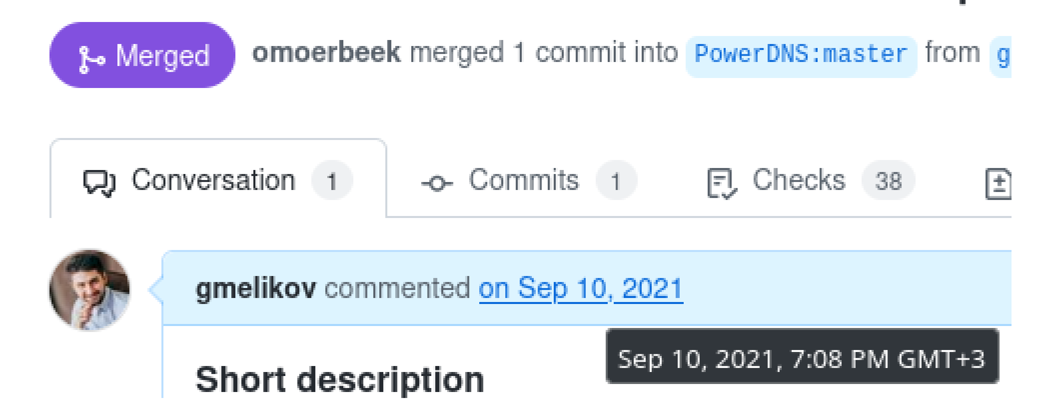
- Не дают менять всё в запросах :(
- Стали контрибьюторами :)
- Storage – уже поздно



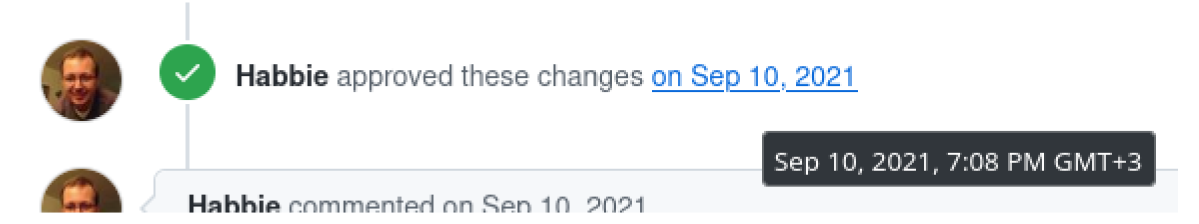


PDNS: Fastest approve in wild-opensource

(для первого пулл-реквеста в проект)



Thank you, @Habbie!





@ https://github.com/PowerDNS/pdns/pull/10714





Другие варианты

- Dnsmasq – на масштабах есть проблемы, нет мультитенантности





Другие варианты

- Dnsmasq – на масштабах есть проблемы, нет мультитенантности
- K8S – CoreDNS
  - Не из коробки
  - Базово тоже без мультитенантности, как и сам K8S





Другие варианты

- Dnsmasq – на масштабах есть проблемы, нет мультитенантности
- K8S – CoreDNS
  - Не из коробки
  - Базово тоже без мультитенантности, как и сам K8S
- Pdnsd – без мультитенантности





Результаты RnD:

- Готового DNS-сервера с мультитенантностью нет





Результаты RnD:

- Готового DNS-сервера с мультитенантностью нет
- Мы не хотим писать ещё один DNS-сервер с нуля





Результаты RnD:

- Готового DNS-сервера с мультитенантностью нет
- Мы не хотим писать ещё один DNS-сервер с нуля
- Добавить честную мультитенантность в существующий public DNS-server дорого





Результаты RnD:

- Готового DNS сервера с мультитенантностью нет
- Мы не хотим писать ещё один DNS сервер с нуля
- Добавить честную мультитенантность в существующий public DNS server дорого
  - Upstream не вольёт такие изменения
  - Поддержка своей downstream ветки – дорого





Что такое “ещё один DNS сервер”

Listing of Vulnerabilities affecting current branches of BIND

#	CVE Number	Short Description
135	2022-38178	Memory leaks in EdDSA DNSSEC verification code
134	2022-38177	Memory leak in ECDSA DNSSEC verification code
133	2022-3080	BIND 9 resolvers configured to answer from stale cache with zero stale-ans
132	2022-2906	Memory leaks in code handling Diffie-Hellman key exchange via TKEY RRs (
131	2022-2881	Buffer overread in statistics channel code
130	2022-2795	Processing large delegations may severely degrade resolver performance
129	2022-1183	Destroying TLS session early triggers assertion failure
128	2022-0667	Assertion failure on delayed DS lookup
127	2022-0635	DNAME insist with synth-from-dnssec enabled
126	2022-0396	DoS from specifically crafted TCP packets
125	2021-25220	DNS forwarders - cache poisoning vulnerability

@ https://kb.isc.org/docs/aa-00913





Какие варианты остаются?





DNS query

```
PROJECT1:~$ dig vk.com a +short
```

```
> Frame 1: 95 bytes on wire (760 bits), 95 bytes captur
> Linux cooked capture v2
> Internet Protocol Version 4, Src: 172.20.107.70, Dst:
> User Datagram Protocol, Src Port: 53679, Dst Port: 53
v Domain Name System (query)
    Transaction ID: 0x89f4
  > Flags: 0x0120 Standard query
    Questions: 1
    Answer RRs: 0
    Authority RRs: 0
    Additional RRs: 1
  v Queries
    > vk.com: type A, class IN
  > Additional records
```

```
08 00 00 00 00 00 00 05  ff fe 04 00 00 00 00 00   ········ ········
00 00 00 00 45 00 00 4b  6f cc 00 00 40 11 07 66   ····E··K o···@··f
ac 14 6b 46 ac 14 40 01  d1 af 00 35 00 37 df f3   ··kF··@· ···5·7··
89 f4 01 20 00 01 00 00  00 00 00 01 02 76 6b 03   ··· ···· ·····vk·
63 6f 6d 00 00 01 00 01  00 00 29 04 d0 00 00 00   com····· ··)·····
00 00 0c 00 0a 00 08 29  bf 12 85 20 1f d4 06      ·······) ··· ···
```

- IPv4/6:
  - source IP
- DNS:
  - Query (domain name)
  - Record types





Идея!

- Просто добавлять суффикс к зонам
  - freelancer.vk.tenant1
  - freelancer.vk.tenant2
- Подойдёт любой DNS server







Путь запроса

1. fl.vk?

host1.vkcs.local.

Private DNS server

магия

host2.vkcs.local.

Upstream DNS Server
1.1.1.1





Путь запроса

1. fl.vk?

host1.vkcs.local.

2. fl.vk → fl.vk**.tenant1**

Private DNS server

магия

host2.vkcs.local.

Upstream DNS Server
1.1.1.1





Путь запроса

1. fl.vk?

host1.vkcs.local.

2. fl.vk → fl.vk.**tenant1**

Private DNS server

3. fl.vk.**tenant1**: 10.0.0.1

магия

host2.vkcs.local.

Upstream DNS Server
1.1.1.1

Путь запроса

1. fl.vk?

host1.vkcs.local.

2. fl.vk → fl.vk**.tenant1**

Private DNS server

3. fl.vk**.tenant1**: 10.0.0.1

магия

4. fl.vk**.tenant1** → fl.vk

host2.vkcs.local.

Upstream DNS Server
1.1.1.1





Путь запроса

1. fl.vk?

host1.vkcs.local.

2. fl.vk → fl.vk**.tenant1**

Private DNS server

3. fl.vk**.tenant1**: 10.0.0.1

магия

4. fl.vk**.tenant1** → fl.vk

host2.vkcs.local.

5. fl.vk: 10.0.0.1

Upstream DNS Server
1.1.1.1

Путь запроса

1. fl.vk?

host1.vkcs.local.

2. fl.vk → fl.vk.**tenant1**

Private DNS server

3. fl.vk.**tenant1**: 10.0.0.1

магия

4. fl.vk.**tenant1** → fl.vk

host2.vkcs.local.

5. fl.vk: 10.0.0.1

Upstream DNS Server
1.1.1.1

3.1. freelance.vk: 1.2.3.4

Нужен proxy

- Rewrite proxy с кастомной логикой
  - т.е. DNS resolver с кастомной логикой







Нужен rewrite proxy (custom DNS resolver)

- dnsdist, by PowerDNS
  - От авторов PowerDNS
  - Те же проблемы с lua plugins







Нужен rewrite proxy (custom DNS resolver)

- Сделаем свой собственнный лёгкий proxy:
  - Изменение query name
  - Изменение records
    - A (IP)





Нужен rewrite proxy (custom DNS resolver)

- Сделаем свой собственнный лёгкий proxy:
  - Изменение query name
  - Изменение records
    - A (IP)
    - PTR





Reverse DNS

```
$dig -x 1.2.3.4
  …
;; QUESTION SECTION:
;4.3.2.1.in-addr.arpa.     IN       PTR

;; ANSWER SECTION:
4.3.2.1.in-addr.arpa. 300 IN       PTR
php1.freelancer.vk
```



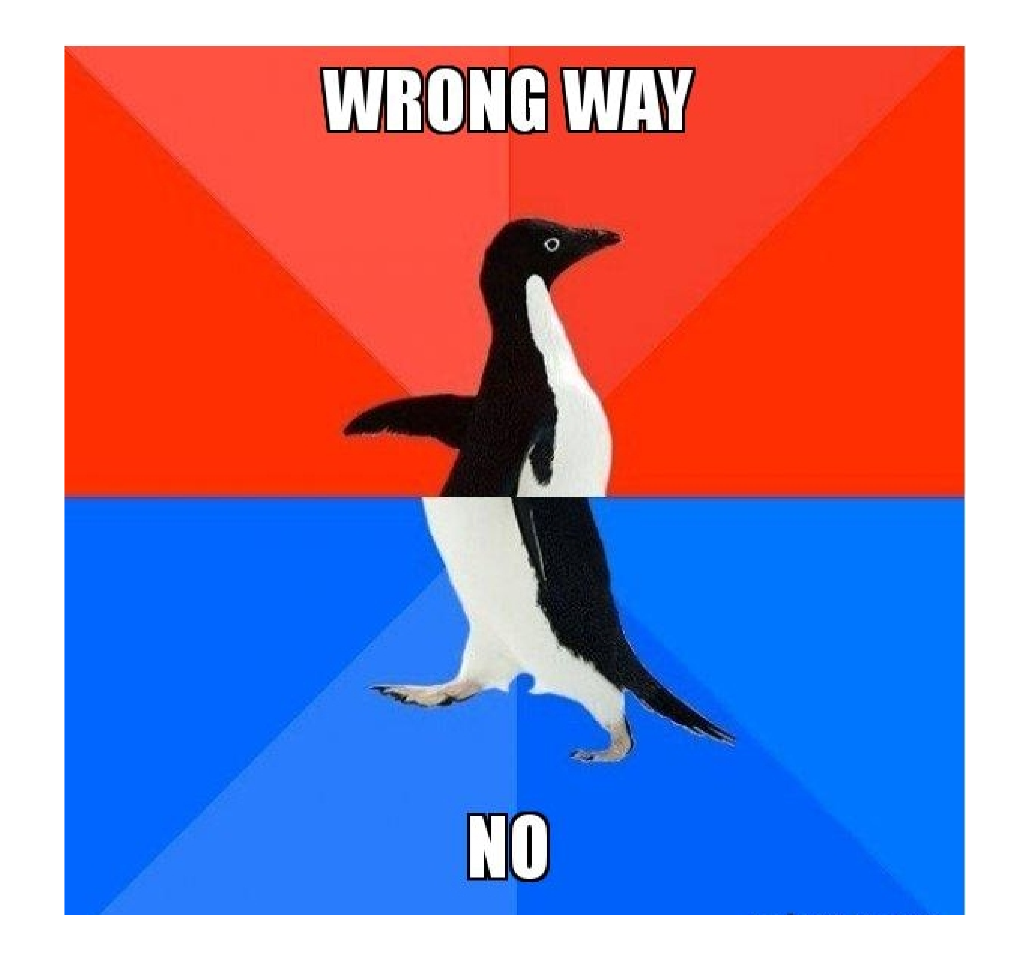




Reverse DNS с суффиксами

```
$dig -x 1.2.3.4
  …
;; QUESTION SECTION:
;4.3.2.1.in-addr.arpa.tenant1    IN
PTR

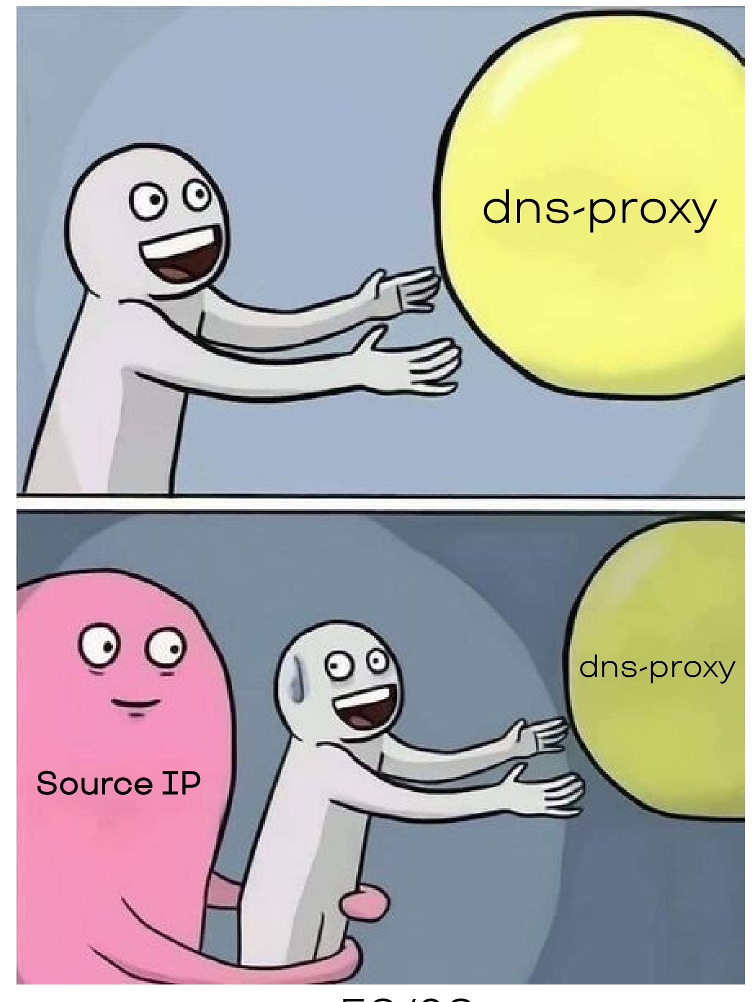
;; ANSWER SECTION:
4.3.2.1.in-addr.arpa.tenant1 300 IN
PTR     php1.freelancer.vk
```







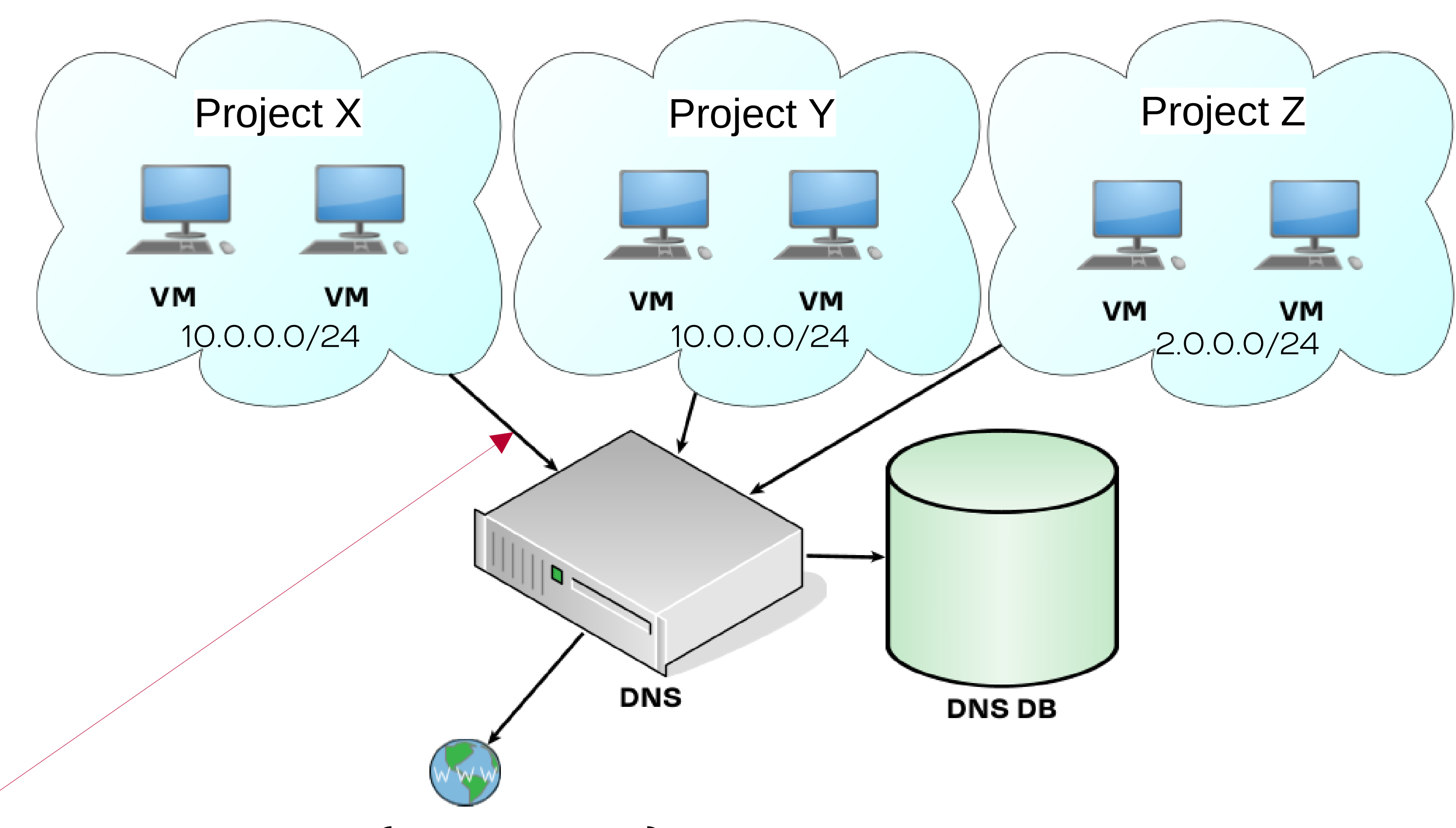
Последняя деталь







Последняя деталь



- VM source IP (private) → tenant ID ?





Последняя деталь

Project X
VM VM
10.0.0.0/24

Project Y
VM VM
10.0.0.0/24

Project Z
VM VM
2.0.0.0/24

SDN

DNS

DNS DB





Software Defined Network?

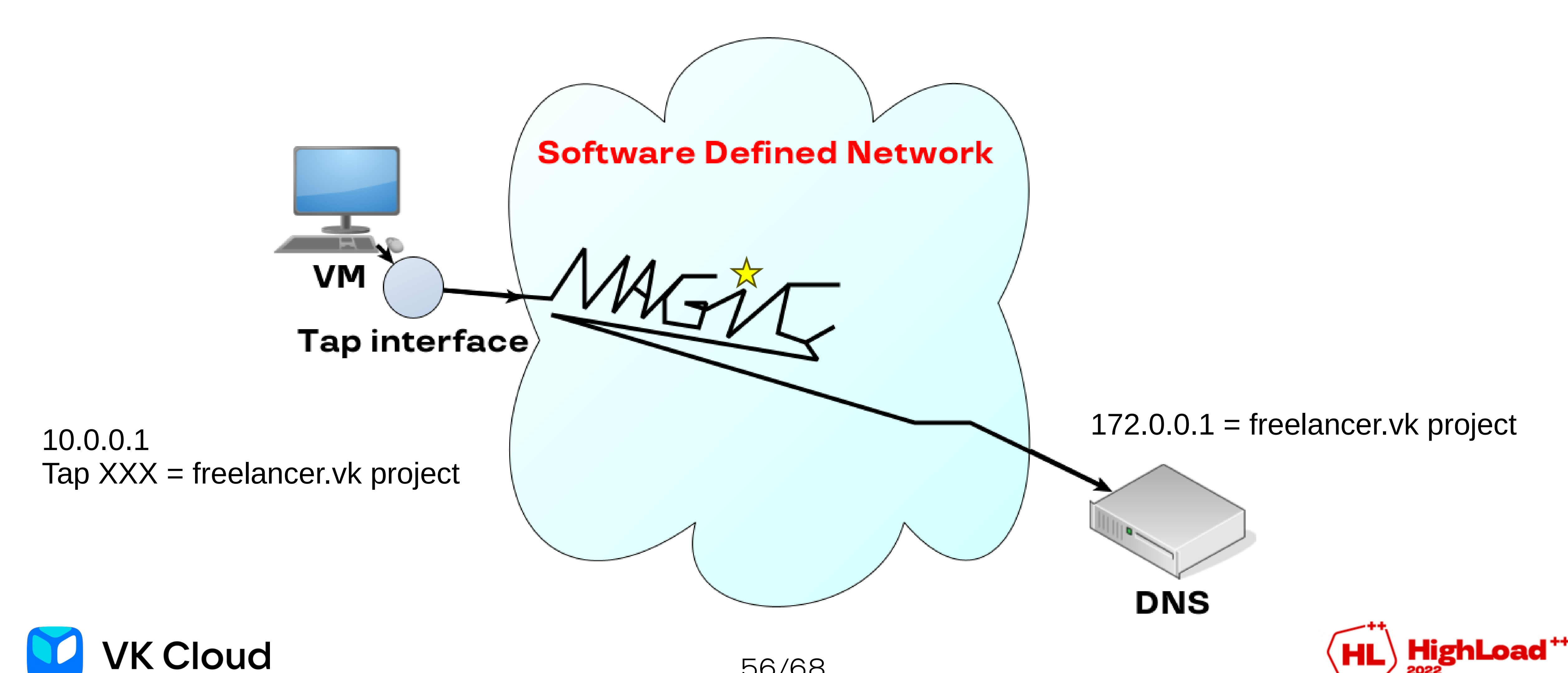
- 100% изоляция тенантов
  - Друг от друга
  - От внутренней инфры
- Обработка и доставка трафика







Последняя деталь

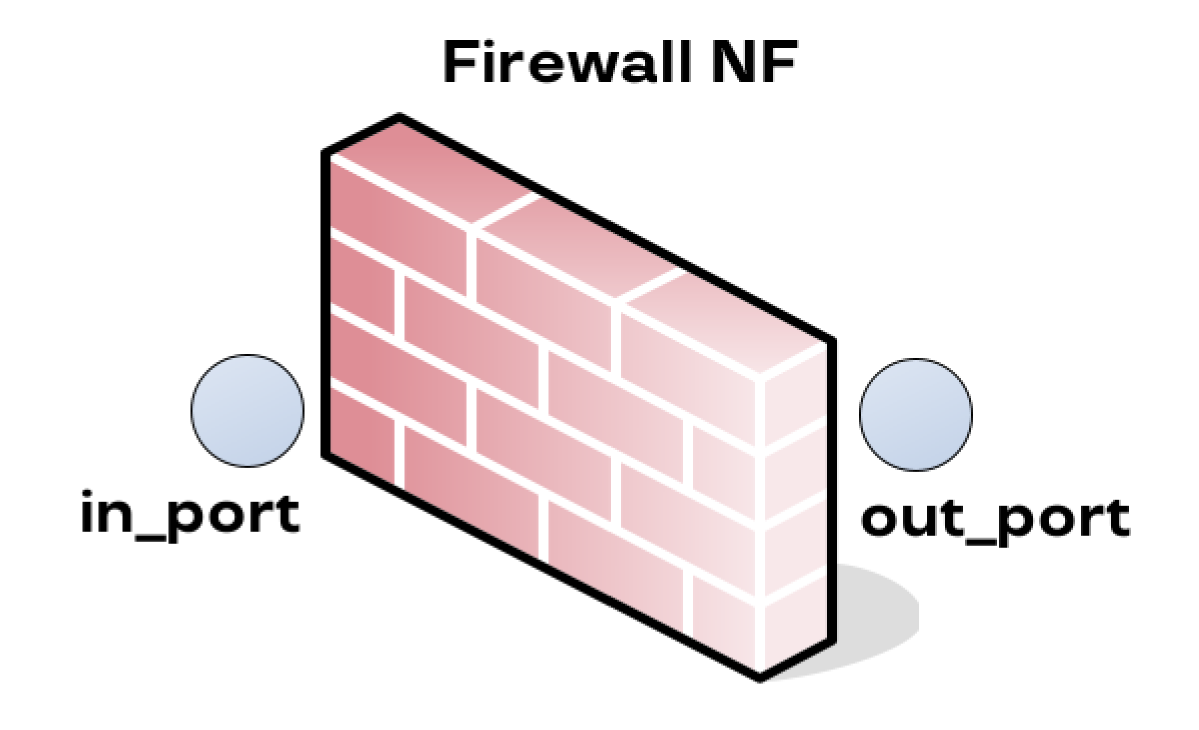


Целых 2 SDN:

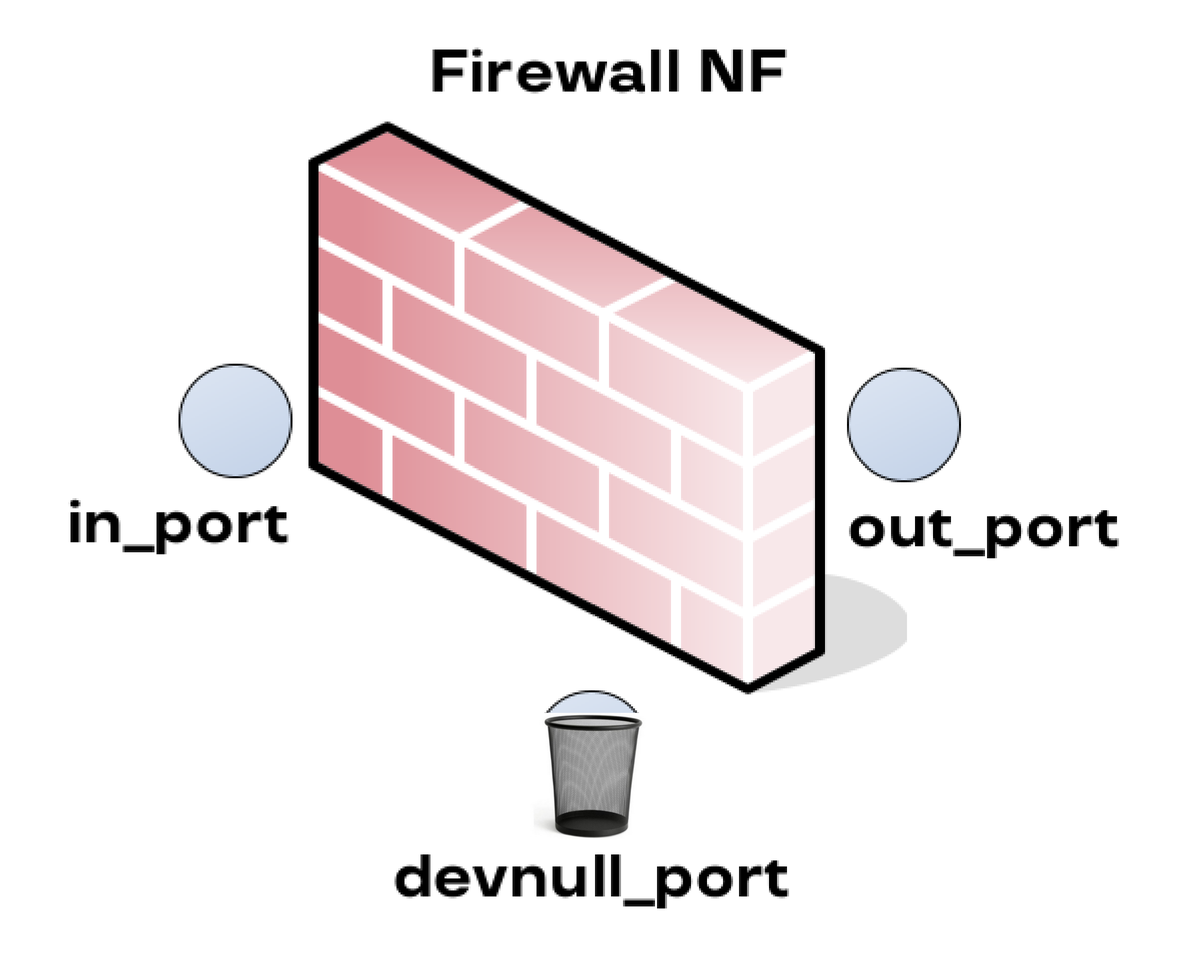




Sprut: network function (NF)



Sprut: network function (NF)







Sprut: network functions (NF)

Service chaining

VM

SG

FIP

Switch

Router

Internet

DHCP

Internet





Sprut: network functions (NF)

VM

FIP

SG

Splitter

Switch

Router

DNS

Internet

DHCP

Internet





Написали DNS-PROXY

host1.vkcs.local.

host2.vkcs.local.

магия

Private DNS server

Upstream DNS Server
1.1.1.1





Написали DNS-PROXY

private network 1

host1.vkcs.local.

dns-proxy

Private DNS server

dns-proxy

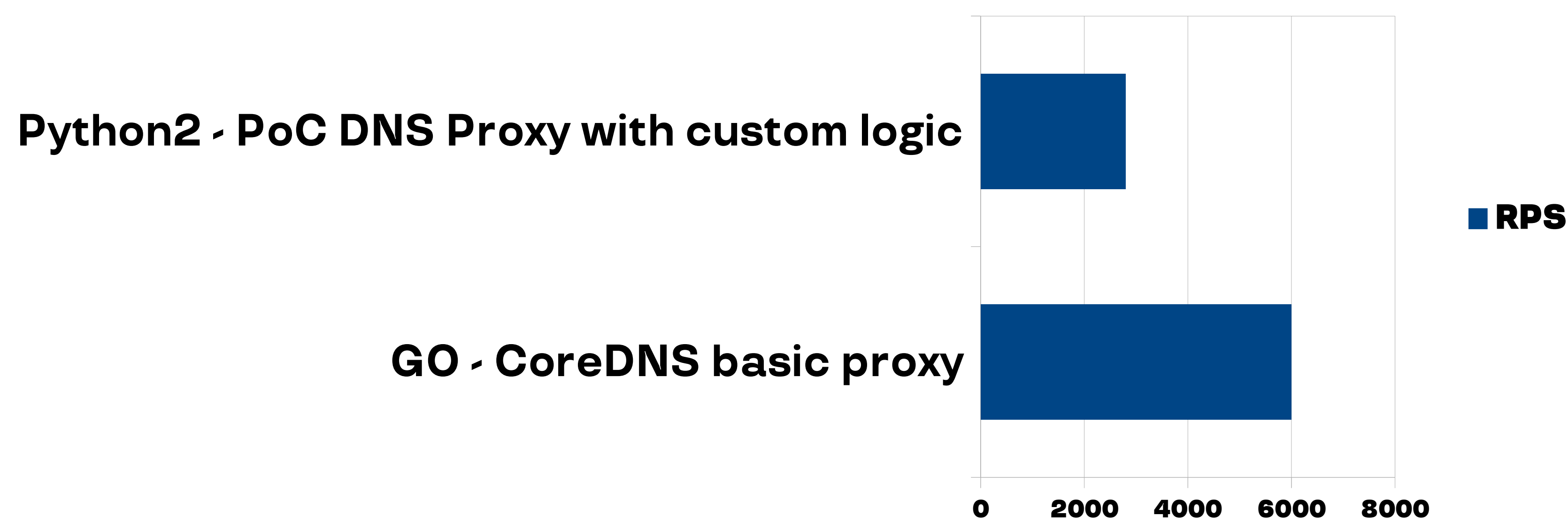
host2.vkcs.local.

private network 2

Upstream DNS Server
1.1.1.1

Приятности

- Даже **python** достаточен!
- От задачи до функционала – 2 человеко-квартала







Выводы

Как добавить мультитенантность во что-то:

- Service с нуля (ещё один DNS auth server)





Выводы

Как добавить мультитенантность во что-то:

- Service с нуля (ещё один DNS auth server)
- Свои патчи в downstream fork





Выводы

Как добавить мультитенантность во что-то:

- Service с нуля (ещё один DNS auth server)
- Свои патчи в downstream fork
- По инстансу на каждый тенант





Выводы

Как добавить мультитенантность во что-то:

- Service с нуля (ещё один DNS auth server)
- Свои патчи в downstream fork
- По инстансу на каждый тенант
- Proxy перед одним сервисом с upstream кодом
  - 100% контроль
  - Нужно поддержать около 1-2% от RFC







Обратная связь
и комментарии по
докладу по ссылке



Меликов Георгий, VK Cloud
@gmelikov

